Earth Surface Processes and Landforms
Earth Surf. Process. Landforms **23,** 1197–1217 (1998)

MAXIMUM ENTROPY SPECTRAL ANALYSIS FOR THE ESTIMATION OF FRACTALS IN TOPOGRAPHY

NICHOLAS J. TATE*

Department of Geography, University of Leicester, Leicester LE1 7RH, UK

Received 22 April 1997; Revised 24 April 1998; Accepted 24 April 1998

ABSTRACT

Spectral analysis is an important method by which the variation in a data set can be decomposed into waves of different frequencies. In the form of the power spectral density it is usually estimated directly from the data using the fast Fourier transform which often requires considerable pre-processing for accurate calculation. Many geomorphological data, including topography, display a power law/fractal model of a decrease in power spectral density with an increase in frequency. Inaccurate calculation of the power spectral density may result in an incorrect estimation of both the power law exponent and observed fractal dimension. As an alternative, maximum entropy spectral analysis provides, in theory, a more accurate estimate of the power spectrum and therefore a more accurate estimate of the fractal dimension. Results are presented of a maximum entropy spectral analysis of both simulated and real topographic surfaces, with derived calculations of fractal dimension. Although the technique offers certain advantages, and returns accurate estimates of the fractal dimension under certain conditions, these have to be traded off against various methodological difficulties which remain unresolved. © 1998 John Wiley & Sons, Ltd.

KEY WORDS spectral analysis; fractal dimension; topography

INTRODUCTION

The analysis of subaerial and submarine topography has revealed patterns of spatial variation which appear to lack a characteristic scale and have been interpreted as evidence of a fractal structure (e.g. Chase, 1992; Klinkenberg and Goodchild, 1992; Herzfeld *et al.*, 1993). However, several problems limit the use of fractal description, particularly the generality of the term fractal, and the plethora of apparently non-coincident estimation methods available to calculate the fractal dimension parameter D (see Klinkenberg (1994) for an overview). The present paper is concerned with a class of fractals strictly referred to as self-affine monofractals to contrast with self-similar fractals and multifractals (see Hastings and Sugihara (1993) and Mandelbrot (1989), respectively, for details). These are typically estimated in the frequency domain using spectral analysis, or in the spatial domain using variogram analysis.

Although the use of spectral analysis is widespread in many areas of geophysics, the use of spatial methods based on the autocorrelation function/variogram are often preferred by many geoscientists for structural analysis and the subsequent calculation of fractal dimension for sequences of data (e.g. Burrough, 1993; Culling, 1986; Robert and Richards, 1988). Indeed, a predisposition *against* the use of spectal analysis in geomorphic investigations has been noted by Hegge and Masselink (1996). This is due largely to the more restrictive statistical stationarity requirements of spectral analysis, and the perception that spectral analysis is inherently complex, with elements of the technique akin to a black art. Arguments for the preferred use of variogram analysis are often made on the basis that it assumes a weaker form of stationarity, as well as being considerably simpler to calculate and often easier to understand. Questions of stationarity are difficult since stationarity *sensu stricto* is a property of the function underlying the spatial data and is not testable from the data themselves (Myers, 1989). Perceived technical complexity can partly be explained by a tendency for statistical texts to concentrate on continuous infinite functions and not the discrete finite data samples obtained from experimental situations. This was noted by Hegge

* Correspondence to: Dr N. Tate, Department of Geography, University of Leicester, Leicester, LE1 7RH, UK
Contract/grant sponsor: NERC

CCC 0197-9337/98/131197–21 $17.50

and Masselink (1996) and their solution was to make a clear description and prescription of the techniques of spectral analysis in various geomorphic contexts. Although this is helpful in explaining and demystifying spectral analysis, their analysis does not consider any of the more advanced approaches to the estimation of spectra, including autoregressive, entropy-based techniques.

Although the use of autoregressive entropy-based spectral analysis has been criticised by some (e.g. Thompson, 1990), evidence suggests (e.g. Fougere, 1985) that it offers real advantages for the analysis of data which follow a power law/fractal frequency structure, where the bulk of variation occurs in the longer wavelengths including topography and many geomorphic data series. The aim of the present paper is to explore the use of autoregressive maximum entropy spectral analysis for the construction of power spectra, and the use of these spectra for the calculation of fractal dimension for both simulated and real topographic surfaces. Comparisons will be made with one of the few published examples of the application of this estimation method for the analysis of topography and calculation of fractal dimension (Gallant *et al.*, 1994).

SPECTRAL ANALYSIS

The description of the statistical variability of a data series in the frequency domain is often termed spectral analysis. There are many different types of spectral analysis ranging from traditional Fourier analysis, described below, to the more recently developed singular-spectrum analysis (Vautard and Ghil, 1989; Vautard *et al.*, 1992). All are based on the use of simple wave forms termed sinusoidal or harmonic waves (Hecht, 1987, p.15). However, purely sinusoidal waves have no actual physical existence (Hecht, 1987, p.254), and anharmonic (non-sinusoidal) and non-periodic waves are encountered in reality. Spectral analysis is based on the concept that such anharmonic/non-periodic data series can be expressed as a sum of periodic waves each in the form of a sinusoidal function defined at a certain frequency f, and amplitude A. The mathematical technique for doing this was devised by the French physicist Jean Baptist Joseph, Baron de Fourier, and has become known as Fourier's theorem. For a function $x(t)$ where t is time, then this theorem states that $x(t)$ having wavelength λ can be synthesized by a sum of harmonic functions whose wavelengths are integral submultiples of λ, i.e. $\lambda/2$, $\lambda/3$ etc. (Hecht, 1987, p.254). This series of periodic components is termed a Fourier series and the process of data transformation which maps $x(t)$ from the space/time domain into the frequency domain $H(f)$ is termed a Fourier transform. A common method of presenting a spectral analysis is in the form of a power spectral density function, which is defined from Parseval's theorem as $1/\lambda|H(f)^2|$ (Maisel, 1971, p.211–213).

The estimation of the power spectral density in this fashion is a standard tool for the analysis of time series data (e.g. Chatfield, 1984, Chapter 4). As such it has been often applied by geoscientists for the analysis of sequences of both temporal and spatial data (e.g. Webster, 1977; Benoist, 1979; Mulla, 1988; Thompson, 1990; Hegge and Masselink, 1996). Furthermore, fitting a linear model to the log–log power spectrum is also one of the more popular methods by which the fractal dimension has been obtained (e.g. Gilbert and Malinverno, 1989; Hough, 1989; Power and Tullis, 1991; Turcotte, 1992, p.77). Although the estimation of the power spectrum is relatively straightforward, it suffers limitations (1) due to the data being sampled at discrete points, or (2) if the series is of a relatively short length. Several alternative methods of estimating the power spectrum are available which seek to overcome these shortcomings (see Robinson (1983) for an overview). Of these, autoregressive maximum entropy spectral analysis, often abbreviated to the acronym MESA or MEM (for maximum entropy method) would appear to be suitable for obtaining the fractal dimension.

Fougere (1985) presents evidence to suggest that MEM spectral analysis provided smoother and higher resolution estimates than the more traditional method of direct calculation using the fast Fourier transform (FFT) for data which posses the bulk of variation at the low frequency part of the spectrum. Such data, characteristic of any geomorphic data sets including the spatial structure of topography, are often referred to as red-noise data, with the longer wavelengths analogous to the red portion of the visible light spectrum. Although estimators based on the FFT are more widely available, there are considerable

problems in obtaining an unbiased, low-varience estimate of the spectrum using this algorithm. An estimate commensurate with the MEM method is only obtained when appropriate detrending and windowing procedures are adopted (Fougere, 1985). The choice of such procedures is often not straightforward. To appreciate the benefits of the MEM approach to spectral estimation, the fundamentals of the FFT estimation method are reviewed below. The reader is referred to Bendat and Piersol (1986), Chatfield (1984) and Press *et al.* (1989) for lucid accounts of both spectral and time series analysis, and Hegge and Masselink (1996) for details of spectral analysis with example applications in geomorphology.

Fast Fourier transform spectral analysis

Before the advent of the FFT algorithm, spectral estimates were often derived indirectly from the discrete Fourier transform (DFT) of the smoothed and truncated sample autocovariance function (Blackman and Tukey, 1958); however, this procedure is largely obsolete (Kanasewich, 1981). The FFT algorithm enables a computationally more efficient calculation of the sample spectrum and operates on the data $x(t)$ which we assume are sampled at N equally spaced points, distance Δt apart to produce a transformed data series:

$$x_n = x(n\Delta t),\ n = 0,\ 1,\ 2,\ ...,\ N-1 \tag{1}$$

Estimation of the sample spectrum $I(f)$ in this fashion is termed periodogram analysis. The first stage in the calculation of $I(f)$ is to calculate the DFT $H(f)$ of x_n which is defined by Bendat and Piersol (1986) as:

$$H(f) = \Delta t \sum_{n=0}^{N-1} x_n \, \mathrm{e}^{2\pi i f n \Delta t} \tag{2}$$

where f is a fundamental frequency f_k of x_n such that $f = k/N\Delta t$ (where k is an integer and $N\Delta t \equiv \lambda$ and $i = \sqrt{-1}$. $H(f)$ is most effectively estimated using the radix 2 FFT algorithm (see Press *et al.* (1989) for details) which requires that N be an integer power of two; data values can be omitted, or padded with zeros if necessary (Bendat and Piersol, 1986). The periodogram estimate is then defined from Parseval's theorem as the modulus squared of Equation 2 to give the power and then divided by $N\Delta t$ to obtain the power spectral density:

$$I(f) = \frac{1}{N\Delta t} |H(f)|^2 \tag{3}$$

Spectral estimates derived from Equation 3 are problematic in that they are prone to high variance and bias (Press *et al.*, 1989). Much of the practical difficulty in estimating the power spectrum arises from the multitude of methods applied to modify $I(f)$ to minimize these problems.

The sole use of $I(f)$ as an estimation of the spectrum is obsolete (Thompson, 1990). The raw periodogram estimates $I(f)$ provide statistically inconsistent estimates of the spectral density: as the data sample size increases, the variability of the spectrum increases such that many spectral peaks may well be spurious (Fougere, 1985; Press *et al.*, 1989). This is a stochastic problem, and some degree of averaging or smoothing is required to reduce the variance of $I(f)$ and produce a consistent spectral estimate (Jenkins and Watts, 1968). There are two main methods of reducing the variance in the periodogram: (1) averaging the N periodogram estimates in the frequency domain (the Daniell estimate) or (2) averaging in the time domain by subdividing the data series into a number of segments, obtaining $I(f)$ for each segment, then

averaging (the Bartlett estimate) (Fougere, 1987; Kanasewich, 1981; Press *et al*., 1989). Unless large data sets are employed (> 2000 points), the Daniell estimate is usually recommended for practical analysis (Kanasewich, 1981). Essentially, this requires the grouping together of adjacent periodogram ordinates $I(f_j)$ in sets of a certain size m, and then the calculation of their average value (Chatfield, 1984; Robinson, 1983):

$$S(f) = \frac{1}{m} \sum_{j=1}^{m} I(f_j) \tag{4}$$

where j varies over m consecutive integers so that the f_j are symmetric about f. As Chatfield (1984) notes, there is little advice within the literature on the selection of parameter m, and he suggests the selection of a value in the region of $N/40$. The trade-off in selecting a larger value of m is between decreased variance, which is desirable, and decreased frequency resolution from a locally averaged spectrum, which is not (Fougere, 1985).

Bias in spectral estimates is caused by the leakage of power from one frequency to another (Press *et al*., 1989, p.422). The use of a simple data set of finite size is, in effect, looking at the data function through a rectangular window or box-car function in the time domain as illustrated in Figure 1A, that is unity inside the measurement interval and zero outside (Fougere, 1985). Because of the very abrupt off–on–off nature of this window, the Fourier transform of the rectangular window leaks power to form side lobes at high frequencies (Thrall, 1983). The Fourier transformation of a finite sample of data is in effect convolving the Fourier transform of the window with the Fourier transform of the data. For example, an infinitely long sinusoid would, when Fourier transformed, produce a single spike (Dirac δ function) at the appropriate frequency, as can be seen in Figure 1B. Truncating the sinusoid and repeating the Fourier transform would introduce a certain breadth into the spectral line, so the spike becomes a lobe, along with the introduction of side lobes, as can be seen in Figure 1C. For a time series with more energy, such side lobes would appear for every Fourier frequency (Kanasewich, 1981) and the true spectrum would be defocused (smeared out) across adjacent frequencies. With red-noise spectra, characterisic of topography, this spectral leakage has the end result of spuriously increasing the power at high frequencies (Bendat and Piersol, 1986; Fougere, 1985; Fox and Hayes, 1985). In fact the rate of decrease of power spectral density with increasing frequency, expected for red-noise data, instead of characterizing the physical process/pattern, would depend on the characteristics of the window alone (Fougere, 1985). There are several methods available for the reduction of bias due to leakage, all of which usually modify x_n in some fashion. The two most widely used methods are data windowing (tapering) and prewhitening.

Data windowing seeks to modify the input data series x_n by de-emphasizing x_n at each end, typically via multiplication with a function that is zero at the end points, and higher towards the centre (Fougere, 1985; Press *et al*., 1989). Combining Equations 2 and 3, the periodogram of the windowed data q_n (termed the modified periodogram) becomes:

$$I(f) = \frac{1}{N\Delta t} \left| \Delta t \sum_{n=0}^{N-1} q_n \, \mathrm{e}^{2\pi i f n \Delta t} \right|^2 \tag{5}$$

which can then be averaged similarly to Equation 4. Much has been written on the subject of optimal window design (see Harris (1978) for details); however, it is often difficult to discriminate between windows, since individual trade-offs may be very subtle (Press *et al*., 1989). At the very least, the use of any non-rectangular window is usually recommended (Fougere, 1985, p.134; Press *et al*., 1989).

The problem of leakage (and the need for windowing) can be generally avoided if the spectrum is flat, for it is the presence of peaks in the spectrum, in combination with the above leakage problem, which spreads power over adjacent frequencies and leads to imprecise spectral estimates (Kanasewich, 1981).

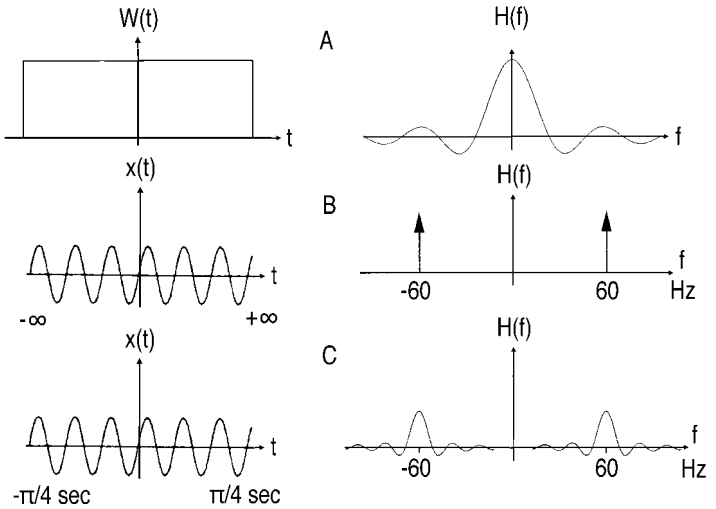

Figure 1. (A) Rectangular window or 'box car' function applied to data in the time domain and its Fourier transform. (B) An infinitely long sinusoid and its Fourier transform. (C) Same as (B) but with a finite data series (modified from Kanasewich, 1981)

The method of prewhitening attempts to flatten the spectrum, by emphasizing those frequencies with lower amplitudes, so that the spectrum is closer to that of white noise (Chatfield, 1984, Chapter 7) and the deleterious effects of the window side lobes are minimized. Essentially this is simply the removal of trend and seasonality in the data, by some type of high-pass filter prior to spectral estimation (Chatfield, 1984, Chapter 7). The procedure is to use the filtered series y_n rather than x_n to estimate the prewhitened spectrum $P(f)$, and then recolour the spectrum to remove the unwanted effects of the filter to estimate $I(f)$. This technique has been adopted in studies of topographic red-noise spectra by Fox and Hayes (1985) and Gilbert and Malinverno (1989). However, a problem arises in the selection of an appropriate filter. Some degree of prewhitening for topography has been obtained by the use of a 'first differencing' filter (e.g. Gilbert and Malinverno, 1989):

$$y_n = x_n - x_{n-1} \tag{6}$$

where x_n is the input, and y_n the filtered output series. As noted by Robinson (1983), prewhitening is often used in conjunction with windowed estimation, rather than as an alternative (Fox and Hayes, 1985; Gilbert and Maliverno, 1989). However, if the filter can be described by an autoregressive model, it is possible to estimate the power spectrum using MEM spectral analysis discussed below.

To avoid the leakage problems caused by having only a finite data sample, the windowing/filtering techniques discussed above are essential. Choosing between the various non-rectangular windows, or choosing an appropriate pre-whitening filter is generally an arbitrary decision. Of greater concern is that these techniques discard or modify a certain part of the data series. Whilst this may be acceptable for digital signal analysis where time series length is not restricted, it is less acceptable for many geomorphic data sets which typically have a limited size. Jaynes' (1982) maximum entropy principle states that in inductive reasoning, our general result should make use of all available information, but at the same time be maximally non-committal about missing information by maximizing the entropy of the underlying probability distribution. This is the basis for the MEM spectral analysis originally developed by Burg (1967, 1968).

MAXIMUM ENTROPY SPECTRAL ANALYSIS

Before describing the method, it is helpful to consider statistical definitions of the terms 'information content' and 'entropy uncertainty'. Ables (1974) forms the basis of the material covered below. Consider a situation in which M different things m_i might happen, the probabilities of occurrence being p_i. If *a priori* all the p_i are equal, then no special information (or degree of surprise) about the system is available. However, if a particular p_i can be determined, a certain amount of information is gained. Therefore the probability of occurrence of an event is directly related to information. Ables (1974) provides the following example. A dice is placed in an opaque box which is shaken and then placed, still closed, on a table. Our information on the state of the dice (i.e. which face is uppermost) is based on what is known about the system; there are six faces, each with a probability p_i such that $0 \leqslant p_i \leqslant 1$ and:

$$\sum_{i=1}^{6} p_i = 1 \tag{7}$$

Although one face must be uppermost, exactly which is of course unknown. However, if the top is removed from the box, the face which is uppermost is revealed, and we gain information about the state of the system. There is a simple measure of this information gain ΔI, best known as Shannon's measure for the entropy uncertainty, or information content, of the probability distribution (Shannon and Weaver, 1949). Three basic axioms allow us to derive the information measure (Thomas, 1981): (1) if $p_i = 1$ then $\Delta I = 0$ since the occurrence of an absolutely certain event tells us nothing we do not already know; consequently (2) ΔI should be a decreasing function of probability; and (3) the information measure of a joint event should be the sum of the information measure of each event. In fact the relation which satisfies these axioms is:

$$\Delta I = \log\left(\frac{1}{p_i}\right) \tag{8}$$

Shannon's entropy is the equivalent measure of information gain for a discrete probability distribution. For a set of M probabilities $\{p_i\}$ this is given by:

$$H_M = \sum_{i=1}^{M} p_i \log\left(\frac{1}{p_i}\right) \tag{9}$$

According to Jaynes' (1982) principle, the prior probability assignment that describes the available information, but is maximally non-committal with regard to the unavailable information, is the one with maximum entropy (Ulrych and Bishop, 1975). Returning to the example of the dice in the box, p_i would be chosen to maximize H_M, subject to the conditions of $0 \leqslant p_i \leqslant 1$ and Equation 9 where $M = 6$. This is $p_i = 1/6$, as expected.

The method of MEM spectral analysis is an approach to estimating the power spectral density $S(f)$ which is consistent with Jaynes' principle: full use is made of the limited data sample, and minimum assumptions are made about the data beyond. Equation 9 can be used, but the interest is not with estimates of p_i but with $S(f)$. Fortunately entropy gain ΔH can be expressed as a function of $S(f)$ such that (Ables, 1974; Thompson, 1990):

$$\Delta H = \int_{-f_c}^{f_c} \log S(f)\, \mathrm{d}f \tag{10}$$

where the limits $\pm f_c$ correspond to the Nyquist critical frequency range. In order to maximize the entropy, relation ΔH must be made as large as possible. The approach then is to choose $S(f)$ according to Equation 10 which is consistent with the known variability of the data and which at the same time maximizes ΔH. To do this, some measure of our sample data which can be related to $S(f)$ must be used. In fact, the autocorrelation function $\phi(k)$, $k = 0, \pm 1, \pm 2...$, may be used for this purpose, which is generally defined as a statistic correlating the value of a discrete process $f(n)$, $n = 0, \pm 1, \pm 2...$, at one time to the value at another time (Maisel, 1971, p.240). From the Wiener–Khintchine theorem $\phi(k)$ and $S(f)$ are related in the form of a Fourier transform pair (Maisel, 1971, p.219). So, the choice of $S(f)$ becomes one which maximizes ΔH with the constraint that the inverse Fourier transform of $S(f)$ yields the autocorrelation function $\phi(k)$ (Ables, 1974; Chen and Stegen, 1974):

$$\int_{-f_c}^{f_c} \log S(f)\, \mathrm{e}^{-2\pi i f k}\, \mathrm{d}f = \phi(k) \tag{11}$$

All that needs to be done is to find a way of estimating $\phi(k)$, not only for those lags which correspond to the data space, but for the infinitely extended function which includes lags beyond the limits of our data set, and then Fourier transforming to obtain $S(f)$. Rearranging and discretizing Equation 11 gives:

$$S(f) = \frac{1}{2f_c} \sum_{k=1}^{L} \phi(k)_{\mathrm{e}}^{\,2\pi i f k \Delta t} \tag{12}$$

To extrapolate $\phi(k)$ beyond the range of the data available, the right-hand side of Equation 12 can be re-expressed and simplified to obtain the equivalence:

$$S(f) = \frac{a_0}{\left| 1 + \sum_{j=1}^{L} a_j \mathrm{e}^{\,2\pi i f k \Delta t} \right|^2} \approx \sum_{k=-L}^{L} \phi(k)_{\mathrm{e}}^{\,2\pi i f k \Delta t} \tag{13}$$

where a_0 is a scalar, and the coefficients a_j, $j = 1, 2, ..., L$ in the denominator of the left-hand side expression are in the form of a vector of the prediction error filter coefficients estimated from the data (Press *et al.*, 1989). The choice of L can be any integer up to N although in practice it is much less than N. The solving of Equation 13 for the coefficients on the left-hand side is done from the known autocorrelations on the right using the following Toeplitz matrix (Press *et al.*, 1989):

$$\begin{bmatrix} \phi_0 & \phi_1 & \phi_2 & \cdots & \phi_L \\ \phi_1 & \phi_0 & \phi_1 & \cdots & \phi_{L-1} \\ \phi_2 & \phi_1 & \phi_0 & \cdots & \phi_{L-2} \\ \cdots & & & & \cdots \\ \phi_L & \phi_{L-1} & \phi_{L-2} & \cdots & \phi_0 \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ a_2 \\ \cdots \\ a_L \end{bmatrix} = \begin{bmatrix} a_0 \\ 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix} \tag{14}$$

In fact the coefficients a_j are very closely related to the autoregressive (AR) parameters α_L which could be estimated to describe an autoregressive process of order L. Using standard notation, x_t ($\equiv x_n$) is regressed on past values of x_t to produce a model of the form:

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \ldots + \alpha_L x_{t-L} + Z_t \tag{15}$$

where Z_t is a noise/error term (see Ulrych and Bishop (1975) for details). Therefore, the task of estimating the MEM power spectrum is twofold: estimating the coefficients, and defining the order/length of the prediction error filter (or the order of the AR model). The choice of filter length is critical as it will influence the degree of detail in the final spectrum. Although the need for windowing/data tapering is avoided, the problem with MEM spectral analysis is that there is no objective way to select the length of the prediction error filter. Press *et al.* (1989) suggest experimentation using different estimates, and Dowse and Ringo (1989) suggest $N/3$ as a maximum. Gallant *et al.* (1994) made use of the MEM spectrum in the context of analysing topographic profiles with a resolution of 20 m, and they found through trial and error that a filter size of $L = 20$ was sufficient for topography.

The size of the prediction error filter (or the order of the AR series) can be determined by minimization of the residual sum of squares R between the estimated filter coefficients and the data series x_t being modelled:

$$R = \sum_{t=1}^{N} (x_t - \hat{a}_1 x_{t-1} - \ldots - \hat{a}_L x_{t-L})^2 \tag{16}$$

However, it is clear that R can be diminished by simply increasing the number of terms on the right-hand side of Equation 16, and a convenient solution is to favour a high order/large filter size. This is clearly unsatisfactory (Webster and McBratney, 1989) and some sort of discrimination procedure is required. Akaike (1969, 1970) suggested a statistic called the final prediction error *FPE* for selecting the order of an AR model which is defined by:

$$FPE = R + 2k\hat{\sigma}^2 \tag{17}$$

where $\hat{\sigma}^2$ is the estimated residual variance of the AR model, k the number of parameters in the model, and R is the residual sum of squares from the model being assessed (Webster and McBratney, 1989). As Ulrych and Bishop (1975) noted, the magnitude of the first term will decrease with increasing order L of the fit, but the second term will increase. The optimum value of L will be that which minimises the *FPE* of Equation 17. Akaike developed expressions for the efficient estimation of the *FPE*, defined as an estimate of the one-step-ahead prediction error variance of an AR model of order L (Hipel, 1981) and estimated by:

$$\hat{FPE} = \hat{\sigma}^2(L)\left(1 + \frac{L+1}{N}\right)\left(1 - \frac{L+1}{N}\right)^2 \tag{18}$$

Akaike suggested that the AR model with the minimum value of *FPE* should be selected for modelling the data series. Some use of the *FPE* has been made for estimating the filter length for MEM spectral analysis by Ulrych and Bishop (1975) and Dowse and Ringo (1989). However a new parameter, the Akaike information criterion (*AIC*), was subsequently introduced (Akaike, 1974) and has largely replaced the *FPE*. This has the form (Hipel, 1981):

$$AIC = -2 \log ML + 2k \tag{19}$$

where *ML* denotes maximum likelihood. Similar to the *FPE*, minimizing the *AIC* is a trade-off between goodness-of-fit and parsimony. In fact, the two parameters are directly related (Ozaki, 1977) such that:

$$N \log FPE = AIC + \text{constant } O(N^{-1}) \tag{20}$$

FFT and MEM comparison

In a landmark paper, Fougere (1985) presented results of an analysis of simulated data displaying differing power law frequency responses, using the two different methods of power spectrum estimation discussed above. Results indicated that MEM spectral estimation was more accurate than the FFT spectrum, particularly for data exhibiting a red-noise structure. As can be seen in Figure 2, results estimating the power law exponent using the FFT spectrum were only coincident with those from the MEM spectrum where various data modification techniques were applied. In addition, Fougere (1985, p.127) argued that much of the superiority of MEM spectral analysis lay in the smoothness of the spectrum, with real changes in spectral shape unobscured by meaningless detail.

Although Thompson (1990, p.614) considered the MEM method insidious and no better than the direct FFT method, and Press *et al.* (1989) recommended cautious usage, the results of Fougere (1985) make the MEM method an attractive choice for the calculation of fractal dimension *D*, particularly since the fractal dimension is directly obtained from the power law exponent of the power spectrum. In the absence of the windowing/data modification techniques discussed above, the FFT-based method may well underpredict this exponent and return an incorrect estimate of the fractal dimension.

CALCULATION OF FRACTAL DIMENSION

The power spectral density $S(f)$ of a fractal can be defined as a power function of frequency f (Schroeder, 1991, p.121):

$$S(f) = cf^{-b} \tag{21}$$

where c is a positive constant, and $b \in (1,3)$. In fact spectra with $b > 0$ constitute what has been termed above as red noise (Shapiro and Ward, 1960). Fractal dimension D is directly obtained from b such that (Voss, 1988):

$$D = D_T + \frac{3-b}{2} \tag{22}$$

where $D_T = 1$ for a transect and $D_T = 2$ for a surface. Log–log transformation linearizes Equation 21:

$$\log S(f) = \log c - b \log (f) \tag{23}$$

and the estimate of b is then usually obtained from a least squares estimate (LSE) regression of $\log S(f)$ on $\log (f)$. Although Mark and Peucker (1978) suggest that LSE regression ought only be used when the goal is establishing a predictive relationship, and *not* as a descriptive model, its use is widespread for estimating fractal dimension (Hastings and Sugihara, 1993, p.83). LSE regression (e.g. Ferguson, 1978) defines the linear relationship between Y and X such that for the ith observed value of Y_i:

$$Y_i = \alpha + \beta X_i + \varepsilon_i \tag{24}$$

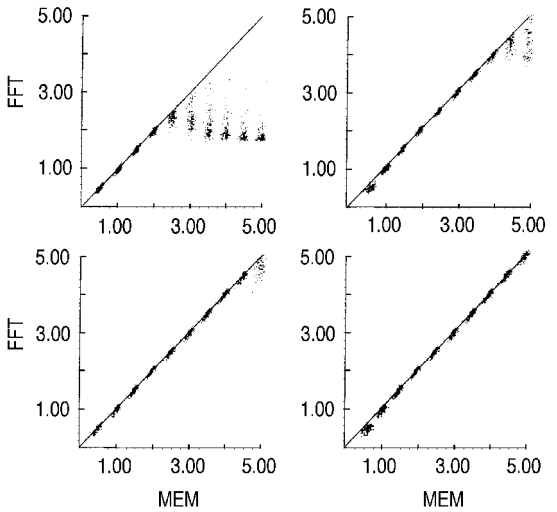

Figure 2. Observed FFT (periodogram) index (where index ≡ negative power) versus MEM index for 100 independent simulations at each index. (A) Raw data; (B) end-matched data; (C) windowed data; (D) end-matched and windowed data (modified from Fougere, 1985)

where ε is a term which incorporates all sources of variability in Y other than X, and β is the LSE estimate of b to give $\beta \in (-1, -3)$. For transect data the estimate of fractal dimension D from the LSE regression simplifies to $2D-5 = \beta$. There are several problems with using LSE regression in this fashion. Points in log–log space are unequally spaced so that log transformation results in the clustering of values at the distal end of the power spectrum. Any fitted regression line is liable to suffer from high leverage values and be unrepresentative of the proximal part of the power spectrum. Solutions to this are to evaluate the power spectrum at a log-frequency interval (e.g. Gallant *et al.*, 1994) or to resample the power spectrum using a geometric progression (to as close an approximation to this as is possible given the frequency interval). The latter will ensure equispacing in the log domain, but at the cost of a dramatic reduction in the number of points used in the line-fitting exercise. Prior to the LSE regression the power spectra calculated below were resampled using a geometric progression to produce between eight and nine values. Another problem is concerned with the assessment of the degree of fit. Traditionally estimates of R^2 have been used; however, there is little advice on what is an acceptable cut-off value. Although, in the context of other methods of estimating fractal dimension D, Andrle (1996) has argued that line-fitting log–log data even with an R^2 in the region of 0·99 (thus implying a linear fit), may contain significant non-linearity. Evans and McClean (1995) considered the choice by Klinkenberg and Goodchild (1992) of 0·90 as a cut-off as too permissive, but do not suggest a better alternative. An alternative to the arbitrary selection of a critical value of R^2 is to estimate confidence limits around the β parameter (and hence the estimate of fractal dimension). From Gardner and Altman (1989) and Wright (1997, p.102) the standard error SE around the slope of β of a regression line may be calculated as:

$$SE(\beta) = \sqrt{\frac{\sum (y-\hat{y})^2 / n-2}{\sum (x-\bar{x})^2}} \tag{25}$$

with 95% confidence limits:

$$\beta - [t_{0\cdot05} \times \mathrm{SE}(\beta)],\ \beta + [t_{0\cdot05} \times \mathrm{SE}(\beta)] \tag{26}$$

where $t_{0\cdot05}$ is the Student's t at $n-2$ degrees of freedom, n equals the number of observations, and $\Sigma\ (y-\hat{y})^2$ is the error (residual) sum of squares. Since D is directly derived from β we can simply adjust Equation 25 to express the confidence limits in terms of fractal dimension.

FRACTAL DIMENSION OF SIMULATED AND REAL TOPOGRAPHY

Attempts to characterize topography by fitting linear functions to estimate the parameter b in Equation 23 pre-date the development of fractals (e.g. Vening-Meinesz, 1951; Bryson and Dutton, 1967; Lettau, 1967; Pike and Rozema, 1975; Sayles and Thomas, 1978; Benoist, 1979). However, with the introduction and development of fractal theory, widespread samples of both subaerial and submarine topography have been analysed and modelled in fractal terms (e.g. Huang and Turcotte, 1989; Klinkenberg and Goodchild, 1992; Herzfeld *et al.*, 1993, 1995; Xu *et al.*, 1993). Indeed, topography defined by fractals appears as a constituent component of several recent dynamic geomorphological landscape models (e.g. Newman and Turcotte, 1990; Sornette and Zhang, 1993; Rigon *et al.*, 1994). The question of whether or not topography is in fact fractal, is peripheral to this paper. However, in spite of the considerable volume of empirical work done, there has been little agreement on the expected magnitude and variability of the observed fractal dimension D for topography, and little comparison has been made between results using different methods of estimating this parameter. In this context, Tate (1995) has compared several methods for the estimation of the variogram and spectral methods for estimating fractal dimension D from a set of topographic surfaces computer-simulated using two-dimensional fractional Brownian motion (fBm) and a set of real 1:24000 30m USGS digital elevation models (DEMs). Simulated surfaces are useful for the testing of estimation methods since they are mathematically well defined with a known autocorrelation structure. Surfaces based on the simulation of fBm are fractal surfaces which can be generated with a known Hurst parameter H and hence fractal dimension D since for a transect across a surface of fBm $D = 2-H$. This parameter can be compared with the observed/calculated fractal dimension D obtained from the application of estimation method. Details of the methods available for the generation of approximate surfaces of fBm can be found in Voss (1988), Musgrave (1993) and Tate (1995). For this research the multidimensional interpolation method proposed by Voss (1985) was programmed from existing pseudocode (Saupe, 1988) to create a set of nine surfaces for each value from $H = 0\cdot1$ to $H = 0\cdot9$.

There are few statistical packages available for the convenient calculation of MEM spectra; the analysis presented here made use of FORTRAN code rewritten from that by Dowse and Ringo (1987) and Press *et al.* (1989). As outlined above, the main difficulty with MEM spectral estimation is choosing the length of the prediction error filter, which determines the degree of detail present in the spectrum. The approach used here is to minimize the Akaike FPE as defined in Equations 17 and 18 with the only constraint that the maximum filter size cannot exceed $N/2$. Whether or not a data series should be detrended prior to spectral estimation when fractal analysis is the aim is not clear from the literature. Bendat and Piersol (1986, p.363) advise caution, suggesting that trend removal should only take place if trends are physically expected or clearly apparent in the data. A solution is to examine the effects of detrending on the resultant spectra and estimates of fractal dimension which will be considered below.

Calculated MEM spectra

Spectra were calculated for each row/column of the sample surfaces described above. An ensemble average for all of the row and the column transects of each surface can be obtained by averaging the power for all rows/columns at each frequency. This average, as well as the range (maximum and minimum) of the log–log transformed power spectral density at each frequency, is displayed at the top of Figure 3 for the row transects of the simulated surfaces. The changing structure of the log–log power

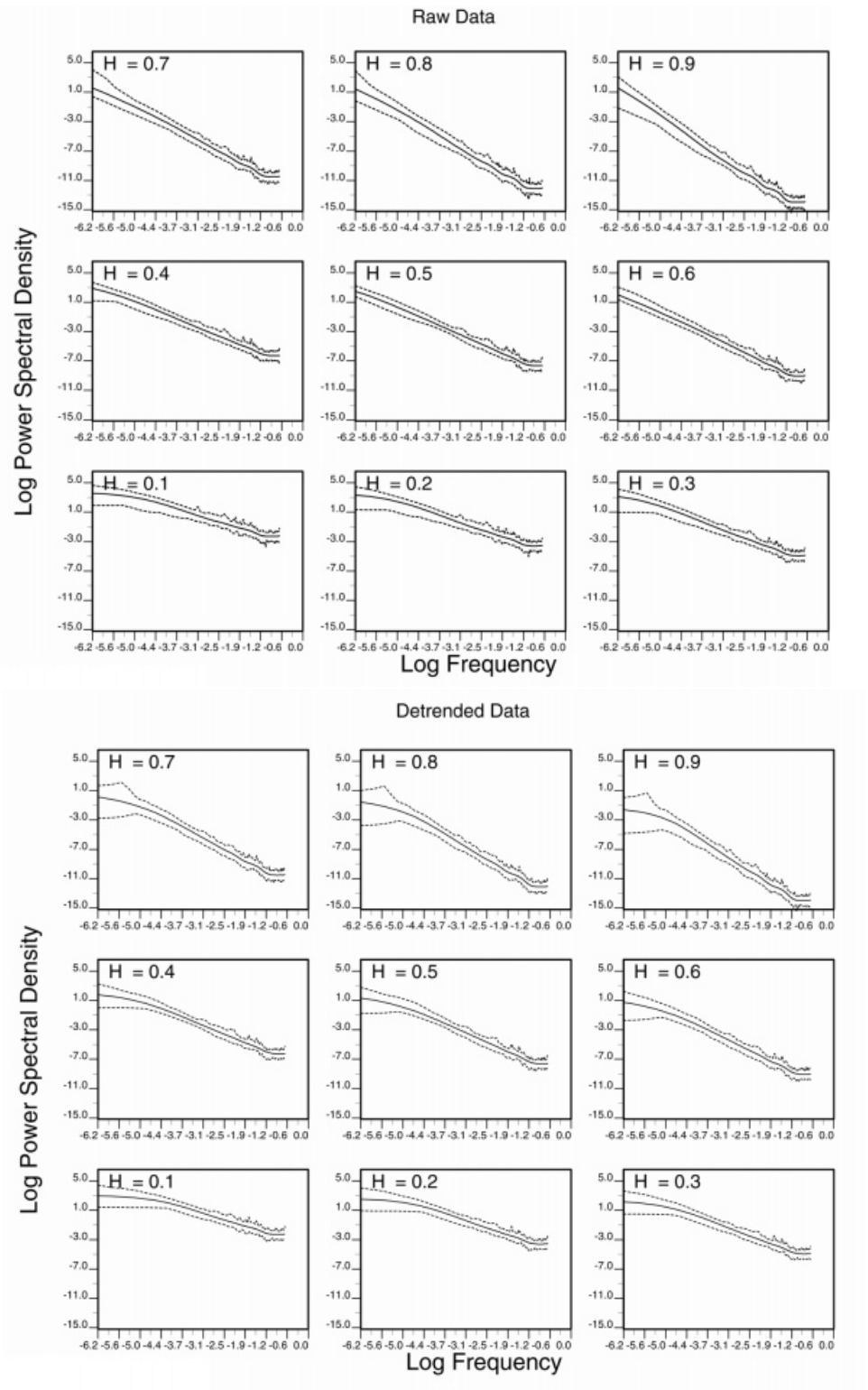

Figure 3. Raw MEM spectra (top) and detrended data (bottom) for the simulated fBm surfaces, showing the mean (solid line) and max./min. (dashed lines) power spectral density at each frequency for row transects

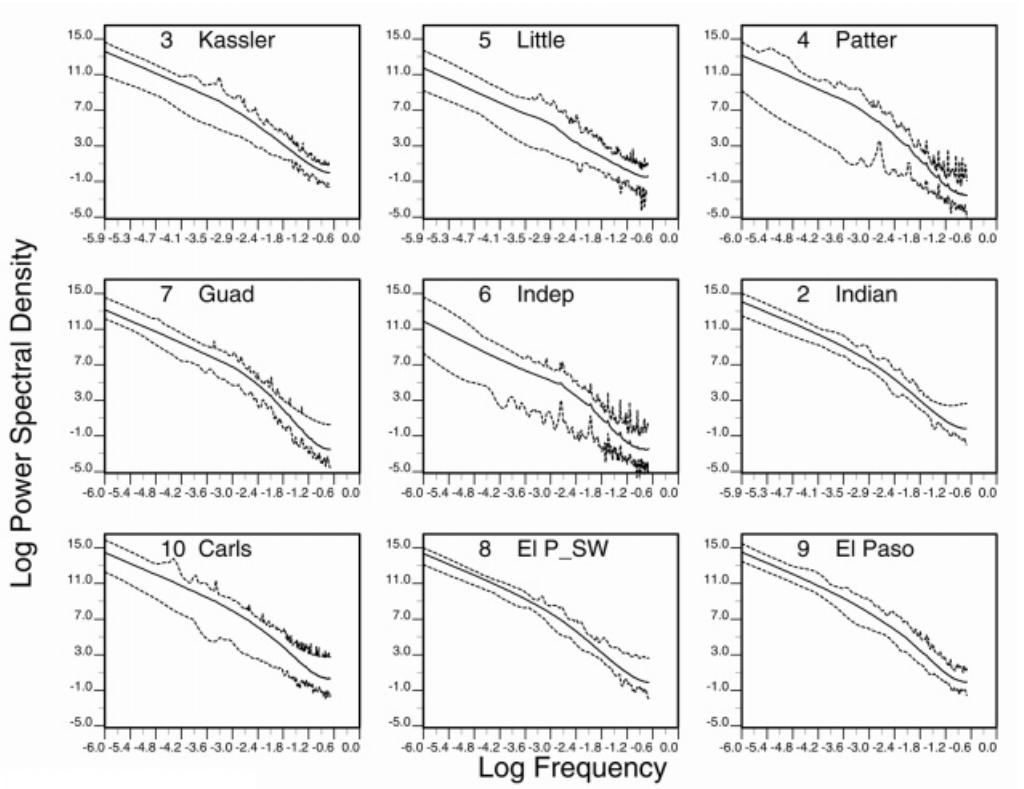

Figure 4. Raw MEM spectra for all DEMs showing the mean (solid line) and max./min. (dashed lines) power spectral density at each frequency

spectrum with parameter H is clearly shown. As expected, the log–log power spectrum is linear with a gradual increase in slope β with increasing H (hence decreasing fractal dimension) as the contribution of the power at high frequencies to the overall spectrum diminishes. Comparative results for the detrended data on removal of a linear trend are displayed at the bottom of Figure 3 where it can be seen that the impact of detrending is most pronounced on the surfaces of largest H (i.e. low fractal dimension). This is entirely reasonable since in contrast to those surfaces of high fractal dimension, these surfaces are composed predominantly of low frequency variation. These results suggest that detrending data in this fashion is unsuitable for calculating fractal dimension, particularly since real topography is largely composed of low frequency variation.

Note that the average spectra flatten off as they approach the cut-off at the Nyquist frequency f_c of 0·5 (log –0·69). This is not easy to explain. With 'real' data the observed pattern would most likely be due to signal aliasing which effectively moves or folds over frequencies which are higher than the Nyquist frequency f_c into the frequency range 0–f_c (see Bendat and Piersol (1986, p.337–339) for details). If this is the case, Press *et al*. (1989) noted that there is effectively little which can be done to remove aliased frequencies other than enforce an arbitrary cut-off by filtering the continuous analogue signal (i.e. prior to obtaining a digitized sample). However, with the simulated data analysed here, there should be no frequencies present higher than the frequency resolution of the original grid which corresponds to f_c. It is therefore more likely that the flattening is because the frequency structure towards the sampling interval is similar to that of white noise (i.e. a lack of structure) and reflects the influence of the random-number generator underlying the fBm simulation method. The portion of the spetrum affected seems to be similar irrespective of the generating value of H.

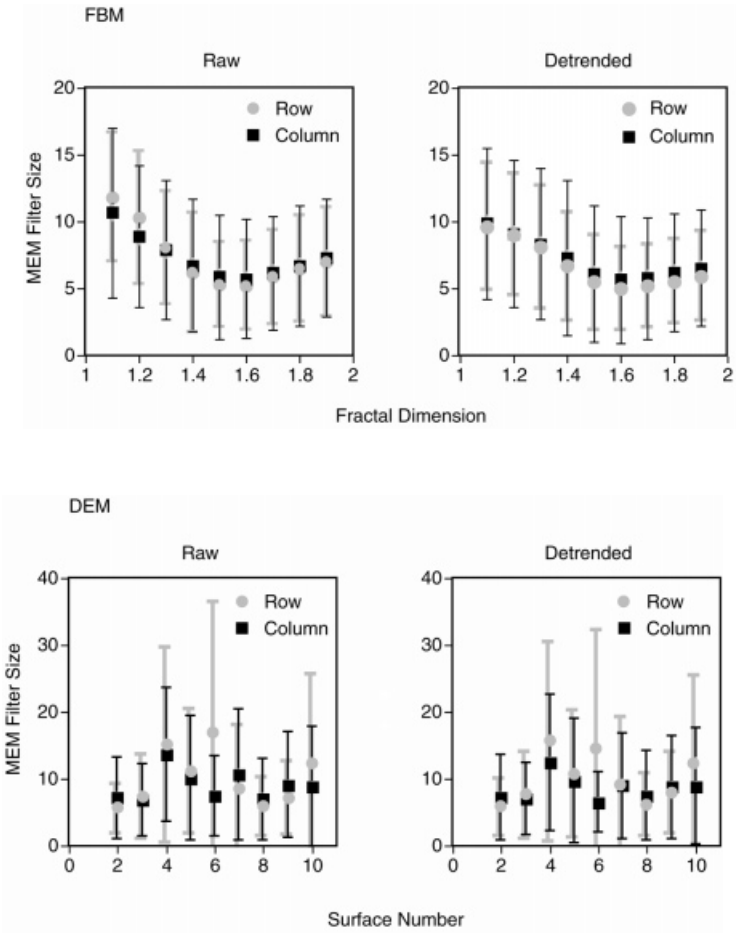

Figure 5. Prediction error filter size for both raw (left) and detrended (right) fBm simulated surfaces (top) and DEM surfaces (bottom)

A similar analysis was performed on a set of nine USGS 1:24000 30m DEMs located in Texas and Colorado, with the averaged spectra displayed in Figure 4. Note that in comparison with Figure 3 the averaged spectra here appear locally more noisy. Although the overall structure of the spectra displayed appears to confirm to the power law/fractal model of Equation 21, there is some suggestion that a single linear model is inappropriate and that these log–log spectra might be better approximated by two separate linear functions or a non-linear (and hence non-fractal) function. Although not displayed here, the results of a similar analysis performed on linearly detrended data support the observation that simple detrending in this fashion is perhaps inappropriate for the calculation of fractal dimension.

Of particular interest with regard to the details of MEM spectral estimation for the simulated surfaces and DEMs is the size of the prediction error filter chosen by the minimization of the Akaike FPE. For the simulated surfaces, the mean and standard deviation of the number of coefficients plotted against the generated fractal dimension for both rows and columns calculated for raw and detrended spectra is displayed in Figure 5 (top). The majority of transects require a filter size not exceeding 17 with the average between 5 and 12 coefficients. The only obvious trend from these graphs is that surfaces of lower fractal dimension as $D \rightarrow 1{\cdot}1$ seem to require a greater number of coefficients. It is not immediately clear as to

why this is so, but it may well reflect the similarity between the size of the prediction error filter and the order of the AR model which was mentioned above.

The mean and standard deviation of the prediction error filter size chosen by minimization of the Akaike FPE for the real surfaces is displayed in Figure 5 (bottom). In comparison with the results for the simulated surfaces described above, there is a much wider range in selected filter size, which reflects the greater variation of real surfaces. Although the mean ranges from 6 to 16 coefficients, the standard deviation ranges from 4 to 19 coefficients. The upper limit corresponds to the spectra calculated for those surfaces which posses both the widest spectral range and the most locally irregular structure (e.g. surfaces 4 and 6).

Calculated fractal dimension

Following the methodology presented above, the power spectra from both the simulated and real surfaces were resampled and then linear functions fitted using LSE regression. For each fitted function, estimates were obtained of β, the fractal dimension D, the standard error SE around this parameter, and the goodness of fit (R^2). The spectra for each surface were analysed in two ways. First, the average ensemble spectra comprising a row and a column power spectrum were used as the input for each LSE regression. This method provides two estimates of fractal dimension for each surface. The second method used was to pass the individual power spectrum calculated for each transect to the regression routine. Whilst computationally more intensive, this method has the advantage that those transects which are poorly fitted by the LSE can be discarded if a critical cut-off for R^2 is selected.

The LSE observed/calculated fractal dimension D_{obs} from the averaged spectra for each surface plotted against the generated/expected fractal dimension D_{gen} of a transect across each surface is displayed in Figure 6. Results are presented for a linear fit of the whole data set, and up to a frequency of log –1·384. Note that the MEM estimator produces estimates of D_{obs} slightly greater than D_{gen} for the range $1{\cdot}2 \leqslant D_{gen} \leqslant 1{\cdot}6$. However, $D_{obs} \approx D_{gen}$ for $D_{gen} > 1{\cdot}6$. Limiting the LSE regression to frequency log –1·384 has the overall effect of decreasing the fractal dimension for all surfaces as displayed in Figure 6. This is not unreasonable given that the effect of the narrowed frequency range will be to fit a curve of steeper slope and therefore reduce D_{obs}. The results would also suggest that for $D_{gen} < 1{\cdot}1$ we find $D_{obs} < D_{gen}$.

Although not presented here, for the detrended spectra using the entire frequency range and trimming the high frequencies, the fractal dimension is overestimated and the confidence limits are wider: the LSE is returning a β which reflects the flattening off of the spectra in the low frequency part of the spectrum. In other words, the spectra are of gentler slope overall, more poorly fit the linear model and hence, from Equation 22, result in a higher fractal dimension. Restricting the LSE estimator to a low frequency cut-off at log –5·32 removes the problem part of the spectrum. The estimates of observed fractal dimension D_{obs} correspond most closely to D_{gen} when both high and low frequencies are excluded from the LSE regression. Confidence limits are wider on the detrended spectra, most likely a reflection of the reduced number of points used in the LSE regression.

It is useful to examine the variation in fractal dimension across each simulated surface. Figure 7 displays results for the raw spectra fitted up to a frequency of log –1·38 for selected simulated surfaces $H = 0{\cdot}2$, $H = 0{\cdot}5$ and $H = 0{\cdot}8$. In comparison with estimates derived from the variogram, Tate (1995) found that the MEM spectral estimator appears to be locally much more noisy, even where the confidence limits are quite narrow as in the middle of Figure 7 for the surface where $H = 0{\cdot}5$. The overall variation in D_{obs} for all three of the surfaces depicted in Figure 7 reflects the summary information shown in Figure 6, indicating that the MEM estimator performs well for $D_{gen} > 1{\cdot}1$, and perhaps most optimally for $D_{gen} \approx 1{\cdot}5$.

Descriptive statistics for the LSE regression for each row and column transect for the simulated surfaces analysed are presented in Figure 8. While estimates of the mean of D_{obs} do not depart greatly from D_{gen}, there is a noticeable spread of values around this mean, particularly for surfaces where

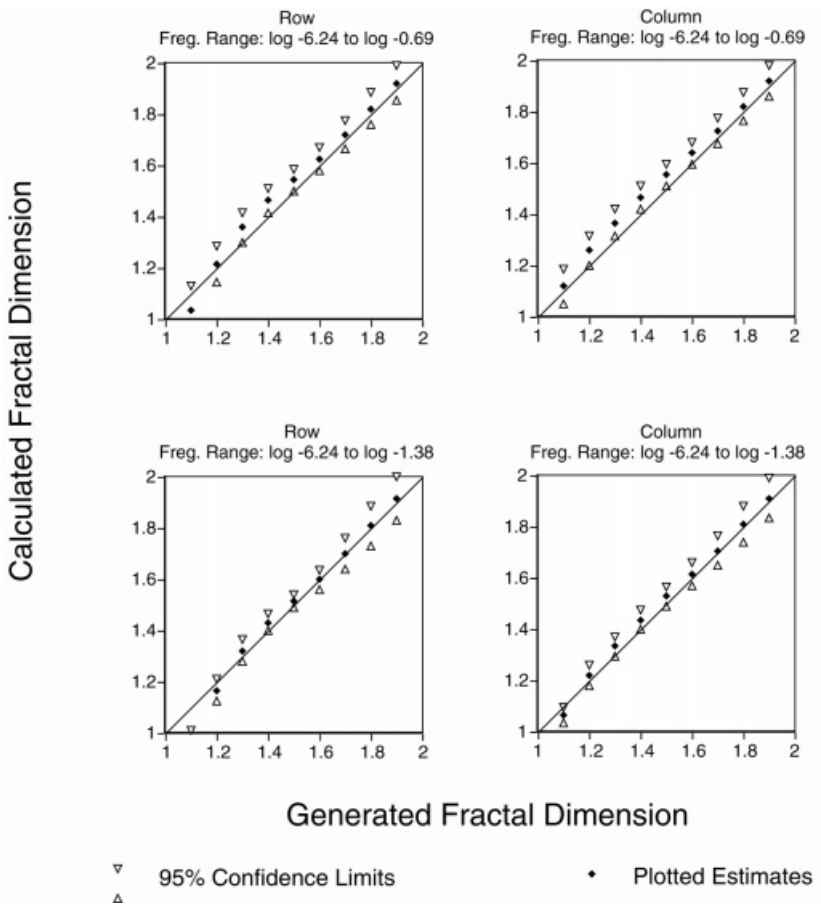

Figure 6. Observed fractal dimension plotted against generated fractal dimension for raw MEM spectra over different frequency ranges: full frequency range (top) and with high frequencies trimmed (bottom)

$D_{gen} < 1{\cdot}3$–$1{\cdot}4$. For both raw and detrended data the width of the standard deviation error bars and the point at which the percentage of the surface where the R^2 parameter exceeds 0·99 is highest around $D_{gen} \approx 1{\cdot}5$. Detrending improves the correspondence of D_{gen} to D_{obs} for the row estimates but has little impact on the column estimates.

For the real topographic surfaces, the estimate of fractal dimensions as each surface is traversed is locally noisy and similar to the results presented in Figure 7. In a similar fashion to Figure 8, it is worth examining the statistics of those transects where the $R^2 > 0{\cdot}99$. The results for fitting over two different frequency ranges are displayed in Figure 9. For the narrower frequency range A, the observed fractal dimension is very variable for both row and column results, with the range exceeding 0·4 in some cases. For the mean estimate the observed range in fractal dimension is approximately $1{\cdot}3 < D < 1{\cdot}5$ and the percentage of the surface fit where $R^2 > 0{\cdot}99$ exceeds 50 per cent in nearly all results. Increasing the frequency range to B produces a pattern of D and R^2 which reflects the convexity of the spectra apparent in Figure 4. Generally across the frequency range, a very low proportion of these surfaces can be described as fractal, and the estimates of D are not representative of the surface as a whole.

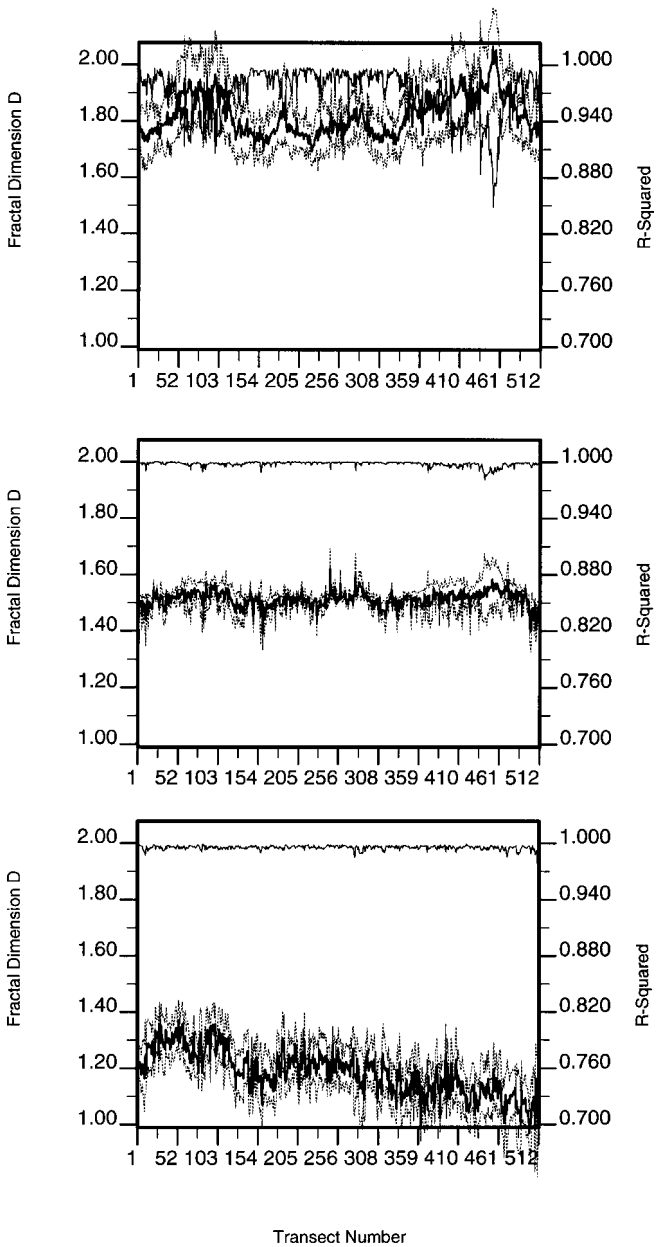

Figure 7. Observed fractal dimension (heavy black line), confidence limits (dashed) and R^2 (light line) fitted for the raw spectra of each row of simulated surfaces $H = 0{\cdot}2$ (top), $H = 0{\cdot}5$ (centre) and $H = 0{\cdot}8$ (bottom)

CONCLUSIONS

In the analysis reported above, MEM spectral analysis performed particularly well in returning the expected fractal dimension D_{gen} for the simulated surfaces. In comparison with the equivalent variogram

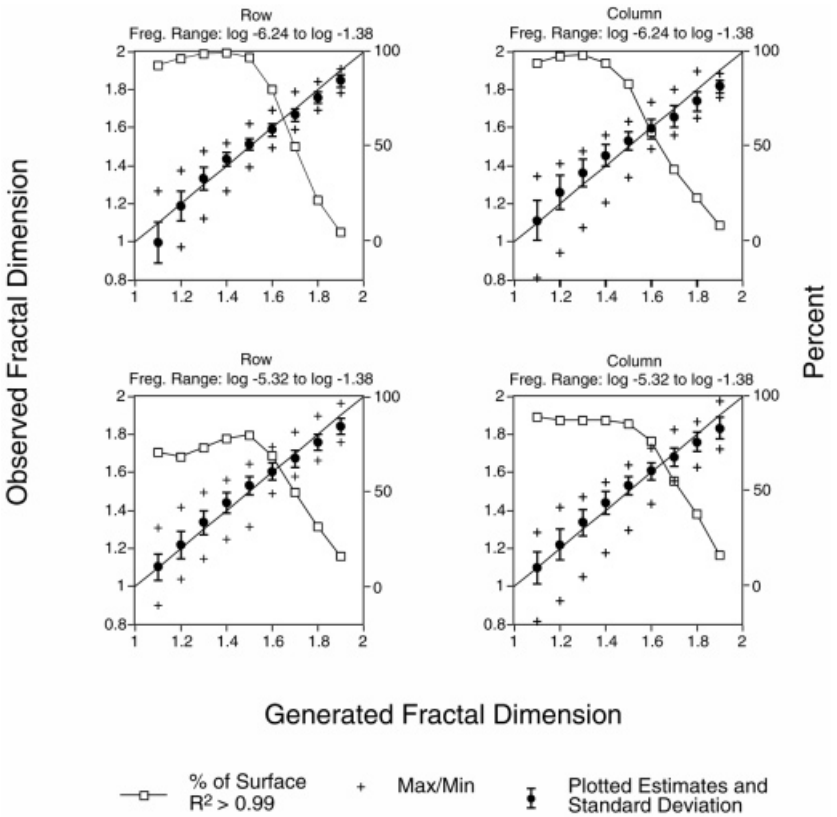

Figure 8. Mean, standard deviation, maximum and minimum values of D_{obs} for those row and column transect spectra for each fBm surface fitted with an $R^2 > 0{\cdot}99$, plotted against the generated fractal dimension D_{gen} for each surface. Also shown is the percentage of each surface where $R^2 > 0{\cdot}99$. spectra fitted to a frequency of log –1·38 (top) and detrended and fitted to a frequency of log –1·38 (bottom)

estimates, Tate (1995) found that the MEM spectral estimator as a whole produces a more accurate estimate for D_{obs} for the range of simulated surfaces considered, particularly as $D \rightarrow 2$. However, as noted above, the MEM spectral estimator is locally quite noisy for all surfaces, particularly the real surfaces. This is not restricted to MEM spectral analysis since similar results were observed by Boger *et al.* (1993) using the more standard method of direct calculation using the FFT.

Although MEM spectral analysis avoids the need to use and experiment with data tapers and windows, the main difficulty is in the 'optimum' selection of the size of the prediction error filter. The approach taken here, minimizing the Akaike FPE, is perhaps more objective than the method of Gallant *et al.* (1994) using a single filter size which represents an *a priori* decision on the variability of data. However, noisy spectra may well be partly due to the variable size of the prediction error filter selected.

Detrending data, which has the result of introducing a more convex structure to the log–log spectra in low frequencies, would seem to be wholly at odds with a fractal model and inappropriate for fractal analysis. Yet, as noted by Myers (1989), the stationarity requirements for spectral analysis usually require that some sort of detrending is undertaken. It is tempting to speculate that some of the fractal results reported here, and perhaps many in the literature, are nothing more than the combination of a non-fractal form with a trend producing an apparent fractal fit. Notwithstanding other difficulties that exist in fractal analysis – that the choice of log–log transformation, whilst standard, may bias the fit of the fractal

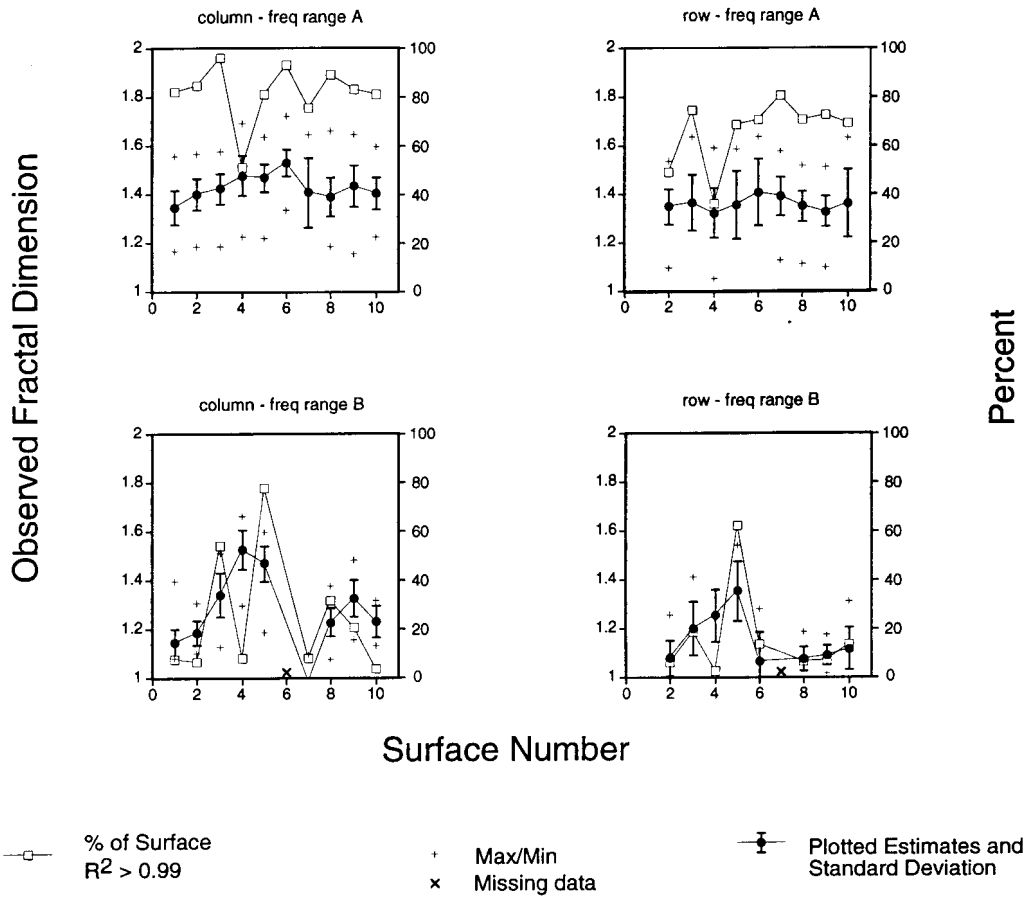

Figure 9. Mean, standard devidation, maximum and minimum values of D_{obs} for those row and column spectra of each DEM surface fitted with an $R^2 > 0{\cdot}99$, plotted against the index number for each surface. Also shown is the percentage of each surface where the $R^2 > 0{\cdot}99$. Frequency range A (top) and B (bottom)

model, with difficulties with the R^2 parameter – this has grave implications for this fractal model of topography as it suggests that a better model might be non-fractal, semi-fractal or even multifractal.

ACKNOWLEDGEMENTS

The work reported here was made possible by a NERC studentship which the author held at the School of Environmental Sciences, University of East Anglia. The author gratefully acknowledges the support of Dr Andrew Lovett, and in particular access to pre-published material from Dr Bob Andrle, Dr John Gallant, Dr Ian Evans, Professor Jens Feder and Dr Brian Klinkenberg, with code/data from Professor Keith Clarke and Dr H. Dowse (c/o MAFF Lowestoft). Figure 1 is from E. W. Kanasewich, Time Sequence Analysis in Geophysics, Third Edition (University of Alberta Press, 1981). Used with permission of the University of Alberta Press. The helpful comments of a referee in improving the paper are also gratefully acknowledged.

REFERENCES

Ables, J. G. 1974. 'Maximum entropy spectral analysis', *Astronomy and Astrophysics* (Suppl), **15**, 383–393.

Akaike, H. 1969. 'Fitting autoregressive models for prediction', *Annals of the Institute of Statistical Mathematics,* **21**, 243–247.

Akaike, H. 1970. 'Statistical predictor identification', *Annals of the Institute of Statistical Mathematics,* **22**, 203–217.

Akaike, H. 1974. 'A new look at statistical model identification', *IEEE Transactions on Automatic Control,* **19**, 716–723.

Andrle, R. 1996. 'The west coast of Britain and statistical self-similarity in nature', *Earth Surface Processes and Landforms,* **21**, 955–962.

Bendat, J. S. and Piersol, A. G. 1986. *Random Data – Analysis and Measurement Procedures*, Wiley and Sons, New York, 566 pp.

Benoist, J. -P. 1979. 'The spectral power density and shadowing function of a glacial microrelief at the decimeter scale', *Journal of Glaciology,* **23**(89), 57–76.

Blackman, R. B. and Tukey, J. W. 1958. *The Measurement of Power Spectra from the Point of View of Communications Engineering*, Dover, New York. 190 pp.

Boger, F., Feder, J. and Jøssang, T. 1993. 'Analysis of 2-dimensional surfaces (unpublished manuscript).

Bryson, R. A. and Dutton, J. A. 1967. 'The variance spectra of certain natural series', in Garrison, W. L. and Marble, D. F. (Eds), *Quantitative Geography Part 2: Physical and Cartographic Topics*, Northwestern University Studies in Geography, 1–14.

Burg, J. P. 1967. 'Maximum entropy spectral analysis', in *Proceedings of the 37th Meeting of the Society of Exploration Geophysicists*, 34–41.

Burg, J. P. 1968. 'A new analysis technique for time series data', paper from *NATO ASI on Signal Processing with Emphasis on Underwater Acoustics*, August 12–23 1968.

Burrough, P. A. 1993. 'Fractals and geostatistical methods in landscape studies, in Lam, N. S-N. and De Cola, L. (Eds), *Fractals in Geography*, Prentice Hall, Englewood Cliffs, New Jersey, 87–121.

Chase, C. G. 1992. 'Fluvial landsculpting and the fractal dimension of topography', *Geomorphology,* **5**, 39–57.

Chatfield, C. 1984. *The Analysis of Time Series – An Introduction*, Chapman and Hall, London, 286 pp.

Chen, W. Y. and Stegen, G. R. 1974. 'Experiments with maximum entropy power spectra of sinusoids', *Journal of Geophysical Research,* **79**(20), 3019–3022.

Culling, W. E. H. 1986. 'Highly erratic spatial variability of soil-pH on Iping Common, West Sussex', *Catena,* **13**(1), 81–98.

Dowse, H. B. and Ringo, J. M. 1989. 'The search for hidden periodicities in biological time series revisited', *Journal of Theoretical Biology,* **139**, 487–515.

Evans, I. S. and McClean, C. J. 1995. 'The land surface is not unifractal: variograms, cirque scale and allometry', *Zeitschrift für Geomorphologie,* **101**(Suppl.), 127–147.

Ferguson, R. 1978. *Linear Regression in Geography*, Concepts and Techniques in Modern Geography No. 15, The Invicta Press, London, 44 pp.

Fougere, P. F. (1985) 'On the accuracy of spectrum analysis of red noise processes using maximum entropy and periodogram methods: simulation studies and application to geophysical data', *Journal of Geophysical Research,* **90,** 4355–4366.

Fox, C. G. and Hayes, D. E. 1985. 'Quantitative methods for analysing the roughness of the seafloor', *Reviews of Geophysics and Space Physics,* **23**(1), 1–48.

Gallant, J. C., Moore, I. D. and Hutchinson, M. F. 1994. 'Estimating the fractal dimension of profiles: A comparison of methods', *Mathematical Geology,* **26**((4), 455–481.

Gardner, M. J. and Altman, D. G. 1989. *Statistics with Confidence*, British Medical Journal, London.

Gilbert, L. E. and Malinverno, A. 1989. 'A characterization of the spectral density of residual ocean floor topography', *Geophysical Research Letters,* **15**(12), 1401–1404.

Harris, F. J. 1978. 'On the use of windows for harmonic analysis with the discrete Fourier transform', *Proceedings IEEE,* **66**, 51–83.

Hastings, H. M. and Sugihara, G. 1993. *Fractals – A Users Guide for the Natural Sciences*, Oxford University Press, Oxford, 235 pp.

Hecht, E. 1987. *Optics*, Addison-Wesley, Reading, Massachusetts, 676 pp.

Hegge, B. J. and Masselink, G. 1996. 'Spectral analysis of geomorphic time series: auto-spectrum', *Earth Surface Processes and Landforms,* **21**, 1021–1040.

Herzfeld, U. C., Kim, I. I., Orcutt, J. A. and Fox, C. G. 1993. 'Fractal geometry and seafloor topography: theoretical concepts versus data analysis for the Juan de Fuca Ridge and the East Pacific Rise', *Annales Geophysicae,* **11**, 532–541.

Hertzfeld, U. C., Kim, I. I. and Orcutt, J. A. 1995. 'Is the ocean floor a fractal?', *Mathematical Geology,* **27**(3), 421–462.

Hipel, K. W. 1981. 'Geophysical model discrimination using the Akaike information criterion', *IEEE Transactions on Automatic Control,* **AC-26**(2), 358–378.

Hough, S. E. 1989. 'On the use of spectral methods for the determination of fractal dimension', *Geophysical Research Letters,* **16**(7), 673–676.

Huang, J. and Turcotte, D. L. 1989. 'Fractal mapping of digitized images: application to the topography of Arizona and comparisons with synthetic images', *Journal of Geophysical Research,* **94**(B6), 7491–7495.

Jaynes, E. T. 1982. 'On the rationale of maximum-entropy methods', *Proceedings IEEE,* **70**, 939–952.

Jenkins, G. M. and Watts, D. G. 1968. *Spectral Analysis and its Applications*, Holden-Day, San Francisco, 525 pp.

Kanasewich, E. R. 1981. *Time Sequence Analysis in Geophysics*, University of Alberta Press, Edmonton, 118–130.

Klinkenberg, B. 1994. 'A review of methods to determine the fractal dimension of linear features', *Mathematical Geology,* **26**(1), 23–46.

Klinkenberg, B. and Goodchild, M. F. 1992. 'The fractal properties of topography: a comparison of methods', *Earth Surface Processes and Landforms,* **17**, 217–234.

Lettau, H. H. 1967. 'Small to large-scale features of boundary layer structure over mountain slopes', *Proceedings, Symposium on Mountain Meteorology*, Atmospheric Science Paper 122, Colorado State University, Fort Collins.

Maisel, L. 1971. *Probability, Statistics and Random Processes*, Simon and Schuster, New York, 280 pp.

Mandelbrot, B. B. 1989. 'Multifractal measures, especially for the geophysicist', *Pure and Applied Geophysics,* **131**(1/2), 5–42.
Mark, D. M. and Peucker, T. K. 1978. 'Regression analysis and geographic models', *Canadian Geographer,* **22**(1), 51–64.
Mulla, D. J. 1988. 'Using geostatistics and spectral analysis to study spatial patterns in the topography of southeastern Washington State, U.S.A.', *Earth Surface Processes and Landforms,* **13**, 389–405.
Musgrave, F. K. 1993. *Methods for Realistic Landscape Imaging*, PhD thesis, Yale University, 268 pp.
Myers, D. E. 1989. 'To be or not to be ... stationary? That is the question', *Mathematical Geology,* **21**(3), 347–362.
Newman, W. I. and Turcotte, D. L. 1990. 'Cascade model for fluvial geomorphology', *Geophysics Journal International,* **100**, 433–439.
Ozaki, T. 1977. 'On the order determination of ARIMA models', *Journal of the Royal Statistical Society Series C (Applied Statistics),* **26**(3), 290–301.
Pike, R. and Rozema, W. 1975. 'Spectral analysis of landforms', *Annals of the Association of American Geographers,* **65**(4), 499–516.
Power, W. L. and Tullis, T. E. 1991. 'Euclidean and fractal models for the description of rock surface roughness', *Journal of Geophysical Research,* **96**(B1), 415–424.
Press, W. H., Flannery, B. P. Teukolsky, S. A. and Vetterling, W. T. 1989. *Numerical Recipes – The Art of Scientific Computing*, Cambridge University Press, 702 pp.
Rigon, R., Rinaldo, A. and Rodriquez-Iturbe, I. 1994. 'On landscape self-organization', *Journal of Geophysical Research,* **99**(B6), 11 971–11 993.
Robert, A. and Richards, K. S. 1988. 'On the modelling of sand bedforms using the semivarigram', *Earth Surface Processes and Landforms,* **13**, 459–473.
Robinson, P. M. 1983. 'Review of various approaches to power spectrum estimation', in Brillinger, D. R. and Krishnaiah, P. R. (Eds), *Handbook of Statistics*, Vol. 3, Elsevier Science Publishers, 343–368.
Saupe, D. 1988. 'Algorithms for random fractals', in Peitgen, H. -O. and Saupe, D. (Eds), *The Science of Fractal Images*, New York, Springer Verlag, 71–136.
Sayles, R. S. and Thomas, T. R. 1978. 'Surface topography as a nonstationary random process', *Nature,* **271**, 431–434.
Schroeder, M. R. 1991. *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise*, W. H. Freeman, New York, 429 pp.
Shannon, C. E. and Weaver, W. 1949. *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, 117 pp.
Shapiro, R. and Ward, F. 1960. 'The time space spectrum of the geostrophic meridional kinetic energy', *Journal of Meteorology,* **17**, 621–626.
Sornette, D. and Zhang, Y.-C. 1993. 'Non-linear langevin model of geomorphic erosion processes', *Geophysics Journal International,* **113**, 382–386.
Tate, N. J. 1995. *The Fractal Dimension of Topography*, PhD thesis, University of East Anglia, 296 pp.
Thomas, R. W. 1981. *Information Statistics in Geography*, Concepts and Techniques in Modern Geography No. 31, Geo Abstracts, Norwich, 42 pp.
Thompson, D. J. 1990. 'Time series analysis of Holocene climatic data', *Philosophical Transactions of the Royal Society of London, Series A,* **330**, 601–616.
Thrall, A. D. 1983. 'Computer programming of spectrum estimation', in Brillinger, D. R. and Krishnaiah, P. R. (Eds), *Handbook of Statistics* **3**, Elsevier Science Publishers, 409–437.
Turcotte, D. L. 1992. *Fractals and Chaos in Geology and Geophysics*, CUP, Cambridge, 221 pp.
Ulrych, T. J. and Bishop, T. N. 1975. 'Maximum entropy spectral analysis and autoregressive decomposition', *Reviews of Geophysics and Space Physics,* **13**(1), 183–200.
Vautard, R. and Ghil, H. 1989. 'Singular spectrum analysis in nonlinear dynamics, with applications to palaeoclimatic time series', *Physica D,* **35**, 395–424.
Vautard, R. Yiou, P. and Ghil, M. 1992. 'Singular-spectrum analysis: a toolkit for short noisy chaotic signals', *Physica D,* **58**, 95–126.
Vening-Meinesz, F. A. 1951. 'A remarkable feature of the earths topography, origin of continents and ocean', *Proceedings, Kosinklijke Nederlandse Akademie Van Wetenshapper: Series B Physical Sciences,* **54**, 212–228.
Voss, R. F. 1985. 'Random fractal forgeries', in Earnshaw, R. A. (Ed.), *Fundamental Algorithms for Computer Graphics*, Springer Verlag, Berlin, 805–835.
Voss, R. F. 1988. 'Fractals in nature: from characterization to simulation', in Pietgen, H. -O. and Saupe, D. (Eds), *The Science of Fractal Images*, Springer-Verlag, New York, 21–69.
Webster, R. 1977. 'Spectral analysis of Gilgai soil', *Australian Journal of Soil Research,* **15**, 191–204.
Webster, R. and McBratney, A. B. 1989. 'On the Akaike Information Criterion for choosing models for variograms of soil properties', *Journal of Soil Science,* **40**, 493–496.
Wright, D. B. 1997. *Understanding Statistics: An Introduction for the Social Sciences*, Sage, London, 228 pp.
Xu, T., Moore, I. D. and Gallant, J. C. 1993. 'Fractals, fractal dimensions and landscapes – a review', *Geomorphology,* **8**, 245–262.

